Climate Model Performance Metrics

Peter J. Gleckler and Karl E. Taylor
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
LLNL

Presentation Outline:

- **Background on model metrics**
- **Exploratory work with simulations from the Coupled Model Intercomparison Project (CMIP)**
  - **Mean climate**
  - **Variability**
- **Cloud-radiative effects**
- **Where do we go from here**
- **The continuing need for new observations**

Climate Model Performance Metrics

Peter J. Gleckler and Karl E. Taylor
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
LLNL

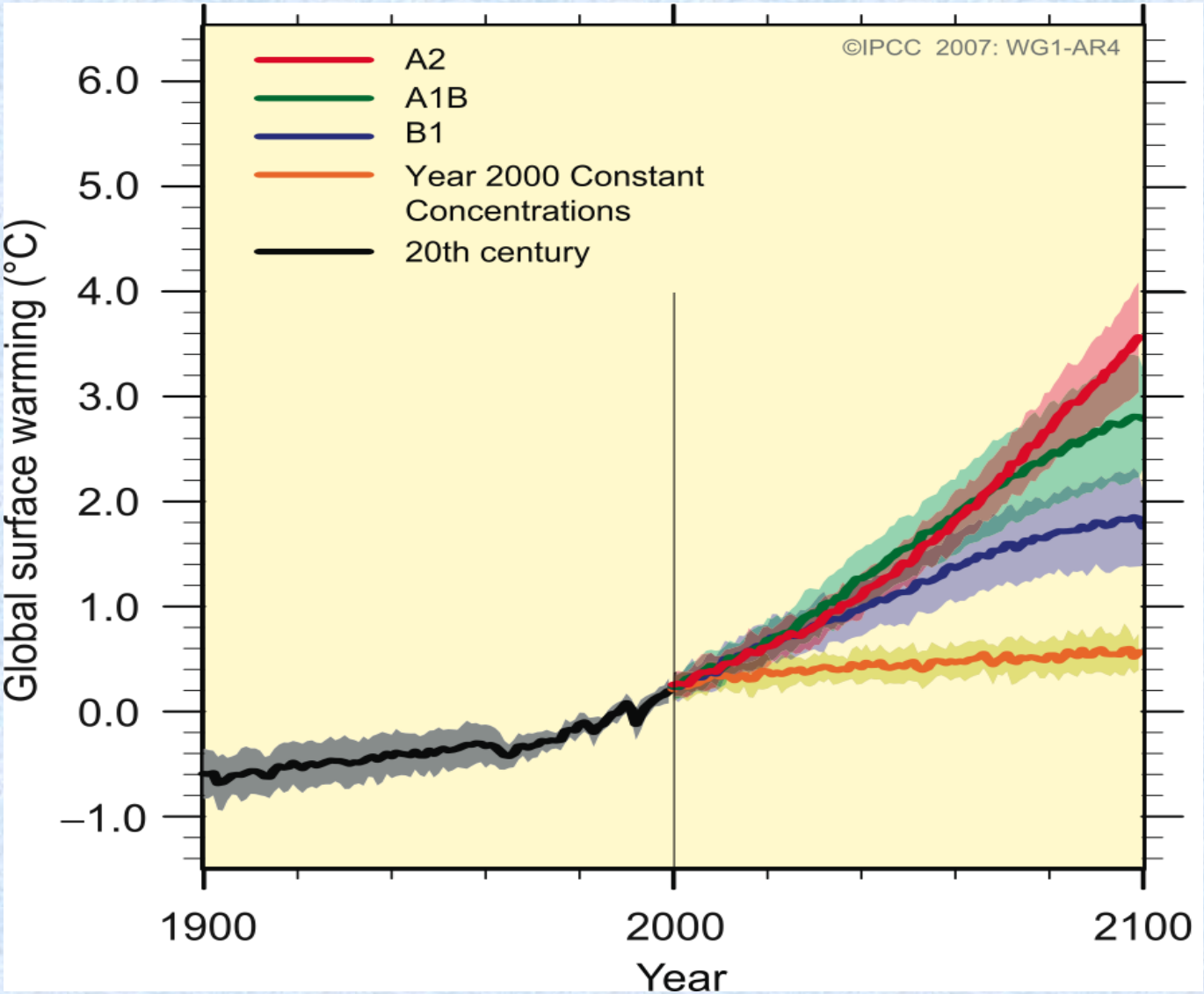
Motivating Questions:

- **Are climate models improving? If so, how rapidly?**
- **Are some models more realistic than others?**
- **How does skill in simulating observed (past and present) climate relate to credibility of model projections?**
- **Can we justify weighting models, based on metrics of skill, to optimize use of multi-model ensembles in making projections of climate change?**

UCRL-ABS-233830

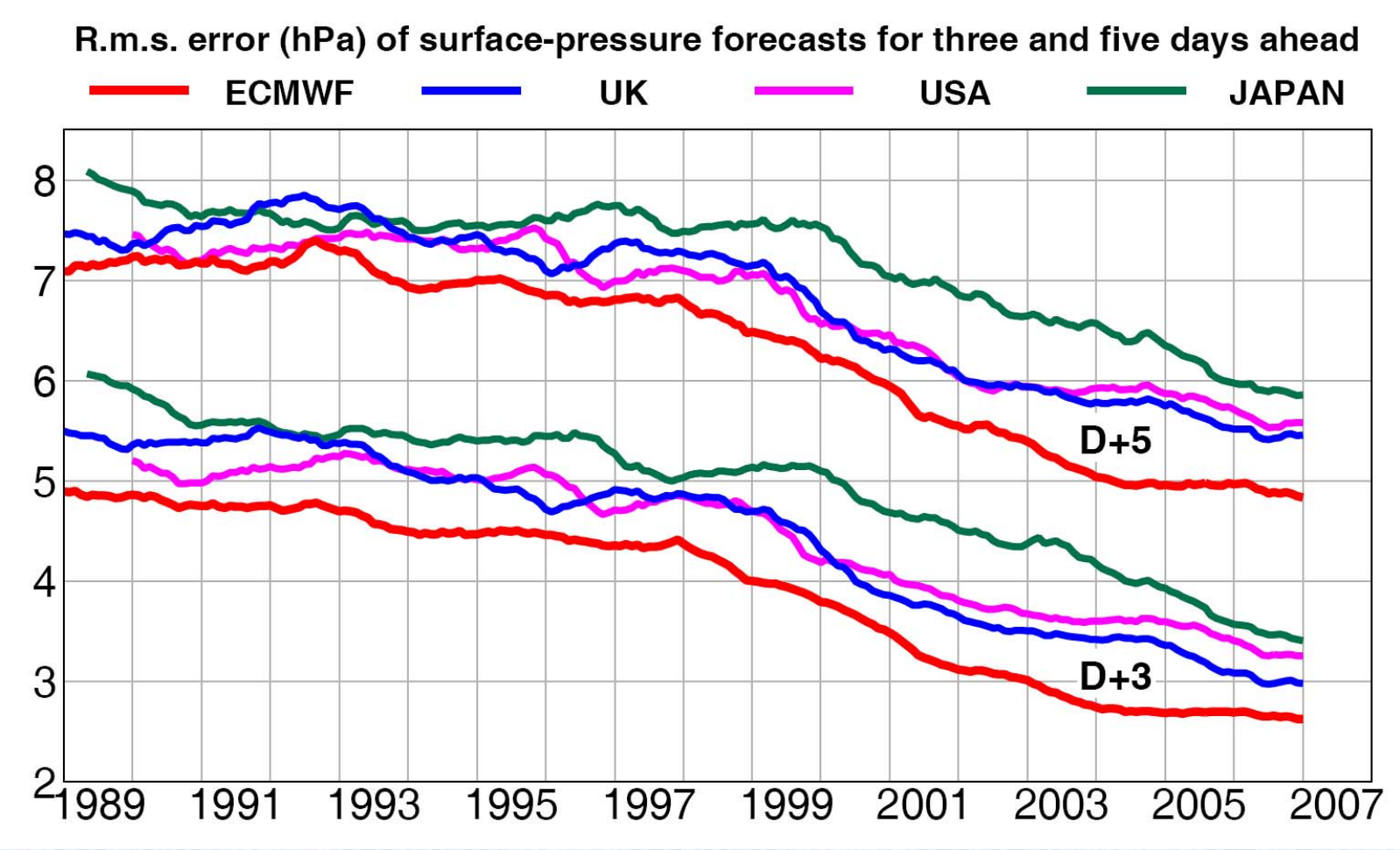
Figure from IPCC AR4 "Summary for Policy Makers"

Global average surface warming as simulated by climate models for different scenarios



Monitoring evolution of model performance: Example from Numerical Weather Prediction

- WGNE routinely reviews still of daily forecasts
- Improvements and deficiencies in the systems identified



Courtesy M.Miller, ECMWF

PCMDI

What do we mean by “metrics”?

- **“Metrics”, as used here, are scalar quantities that objectively measure the quality of a model simulation, e.g.,**
  - → **Skill in simulating things we have observed (“performance metrics”)**
  - → **Model reliability for applications (e.g., “projection reliability metrics”)**
    - **How accurate are model projections of climate change?**
    - **Extremely valuable… and… extremely difficult**
- **Quantify errors, but usually *not* designed to diagnose reasons for model errors**

PCMDI

Some recent work on climate model "performance metrics"

- Gleckler, P., K. Taylor, and C. Doutriaux, 2008:
  Performance metrics for climate models, JGR, *in press*

- Pincus, R., Batstone, C., Hoffman, R., K. Taylor, and P. Gleckler, 2008:
  Evaluating the present-day simulation of clouds, precipitation and radiation in climate models, JGR, accepted

- Reichler, T., Kim J., 2008:
  How well do coupled models simulate today's climate?, BAMS, *in press*

- Williams, K., and M. Webb, 2008:
  A quantitative climate performance assessment of cloud regimes in GCMs, Climate Dynamics, *submitted*

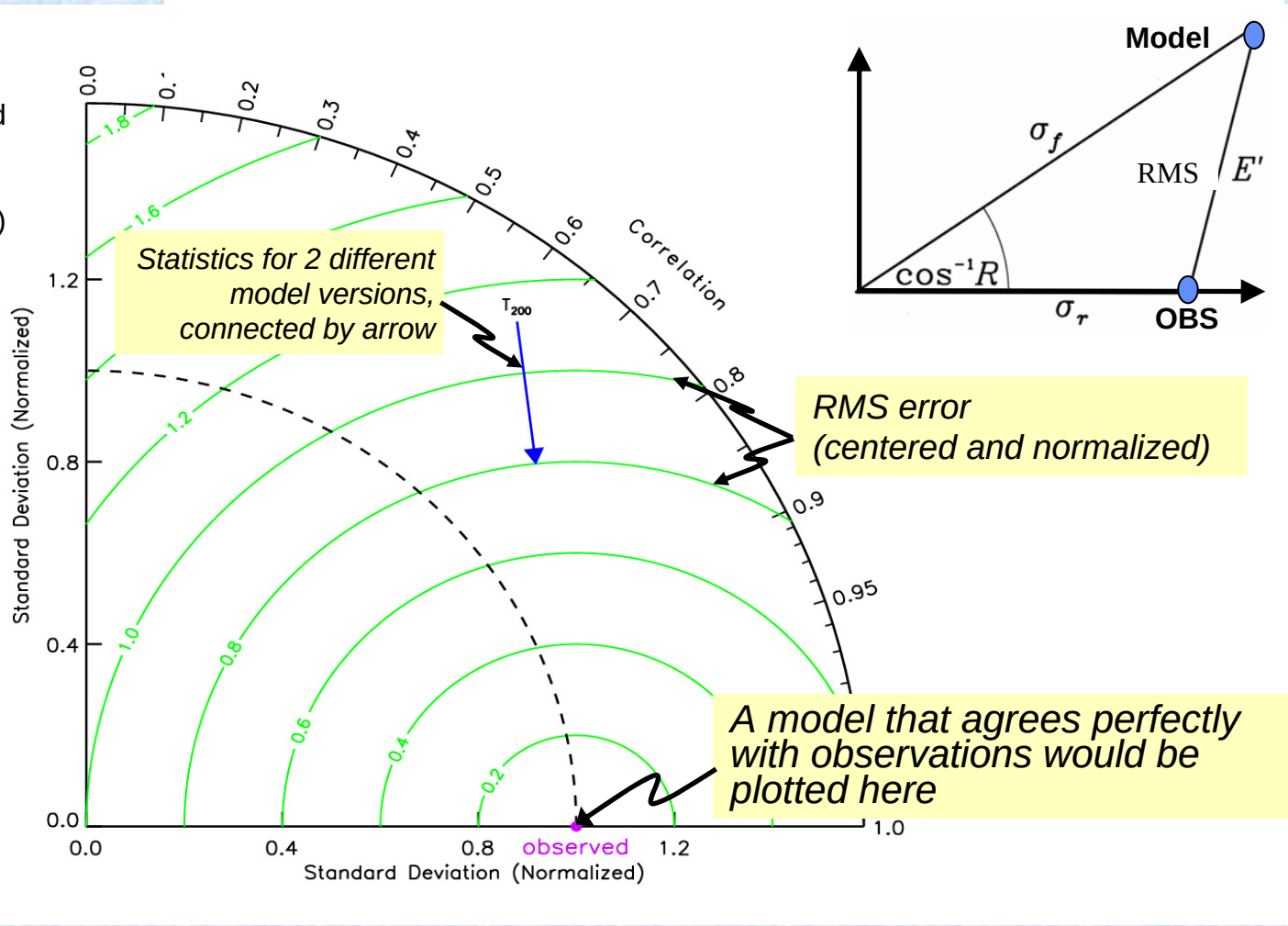
What opportunities are there to evaluate models and build confidence in model physics & dynamics?

- **Model's externally "forced" responses on a range of time-scales:**
  - → **Diurnal cycle**
  - → **Annual cycle**
  - → **Volcanic eruptions, changes in solar irradiance, …**
- **Model's "unforced" behavior (weather, MJO, ENSO, NAO, PDO …)**
- **Evaluate model representation of individual processes and co-variability relationships**
- **Test model ability to solve the "initial value" problem**

PCMDI

Three statistics characterizing agreement between simulated and observed fields can be shown: Taylor Diagram

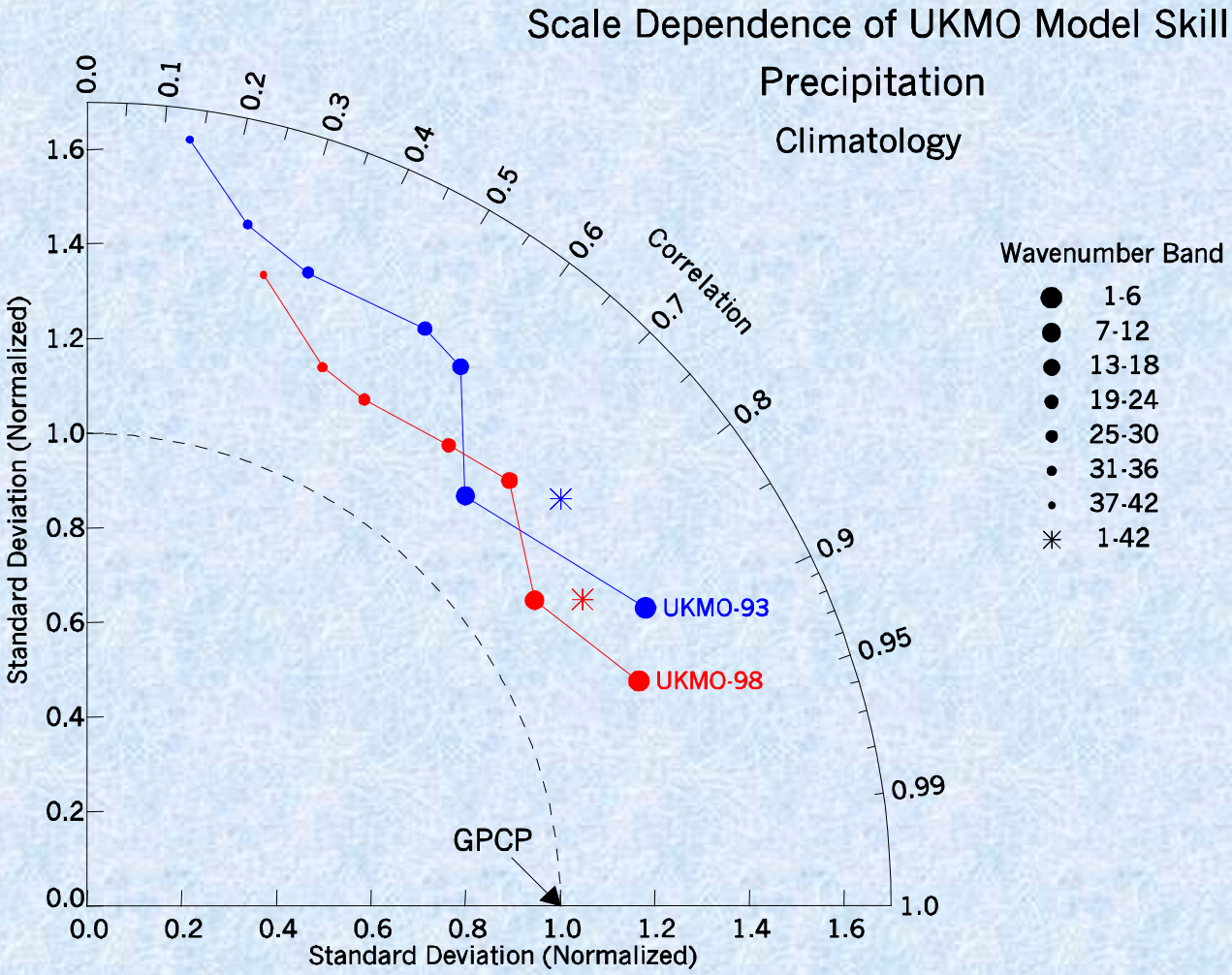
(Note: Can be extended to include negative correlations)



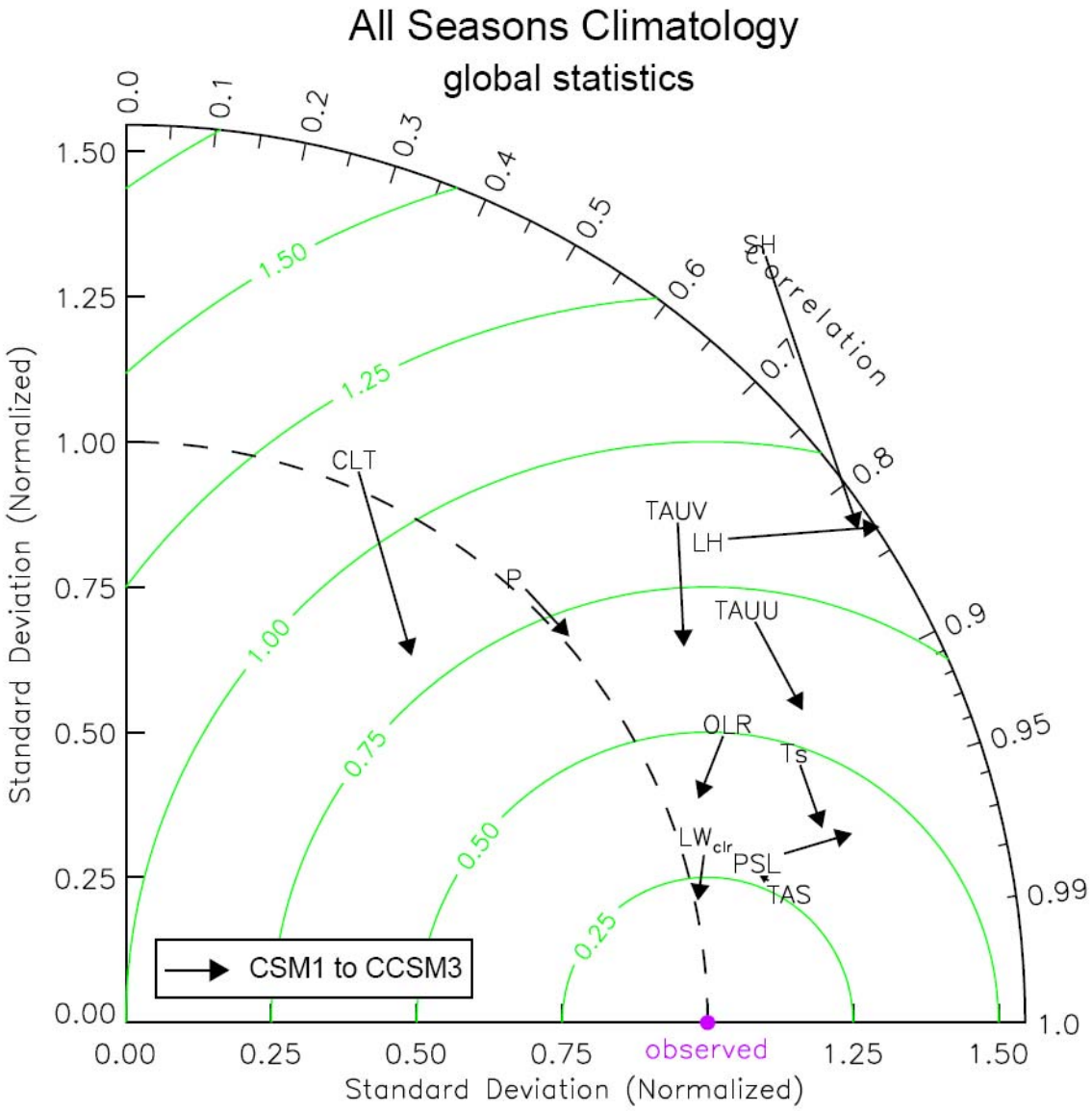
Taylor, *J. Geophys. Res.* (2001)

PCMDI

The larger the scale the better the model skill



Tracking model performance in the development process



Providing feedback to NCAR on newer model versions

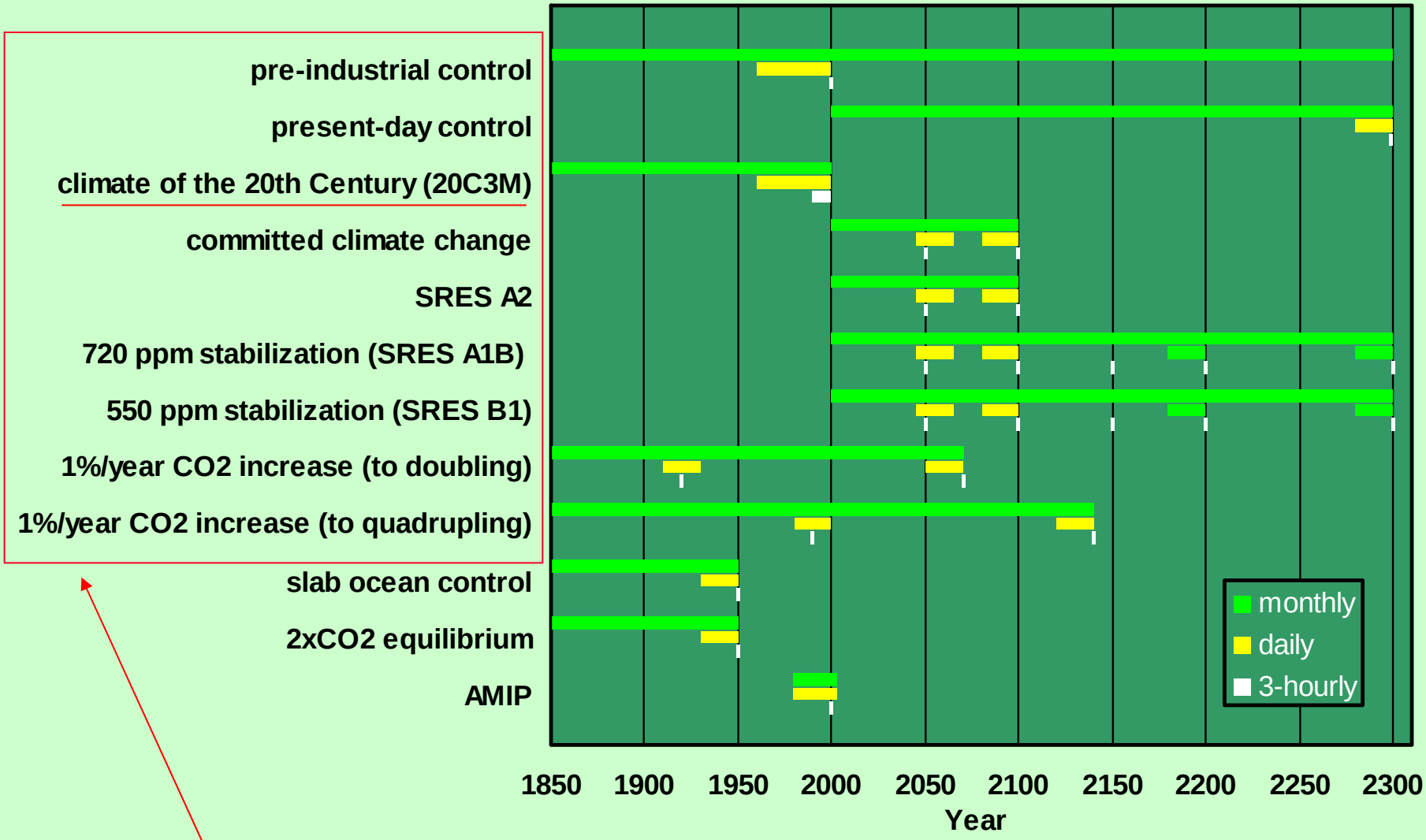
PCMDI

The CMIP3 multi-model dataset

- 2003-2004: In anticipation of the IPCC AR4, PCMDI assisted the World Climate Research Programme's Working Group on Coupled Modelling (WGCM) in the design and coordination a new suite of experiments
- 2004-2005: Modeling groups performed simulations and submitted standardized output to PCMDI for dissemination
- 2005-present: Early publications form the basis of model analysis in the IPCC AR4. To date, over 250 publications based on CMIP3

PCMDI

Sampling by experiment



Fully coupled OAGCM runs

External forcings applied in the "20th Century" simulations

	Model	G	O	SD	SI	BC	OC	MD	SS	LU	SO	V
1	CCCma-CGCM3.1(T47)											
2	CCSM3											
3	CNRM-CM3											
4	CSIRO-Mk3.0											
5	ECHAM5/MPI-OM											
6	FGOALS-g1.0											
7	GFDL-CM2.0											
8	GFDL-CM2.1											
9	GISS-AOM											
10	GISS-EH											
11	GISS-ER											
12	INM-CM3.0											
13	IPSL-CM4											
14	MIROC3.2(medres)											
15	MIROC3.2(hires)											
16	MRI-CGCM2.3.2											
17	PCM											
18	UKMO-HadCM3											
19	UKMO-HadGEM1											

Well-mixed GHGs | Ozone | Sulfate (direct) | Sulfate (indirect) | Black carbon | Organic carbon

Mineral dust | Sea salt | Land use | Solar irradiance | Volcanic aerosols

Reference data sets

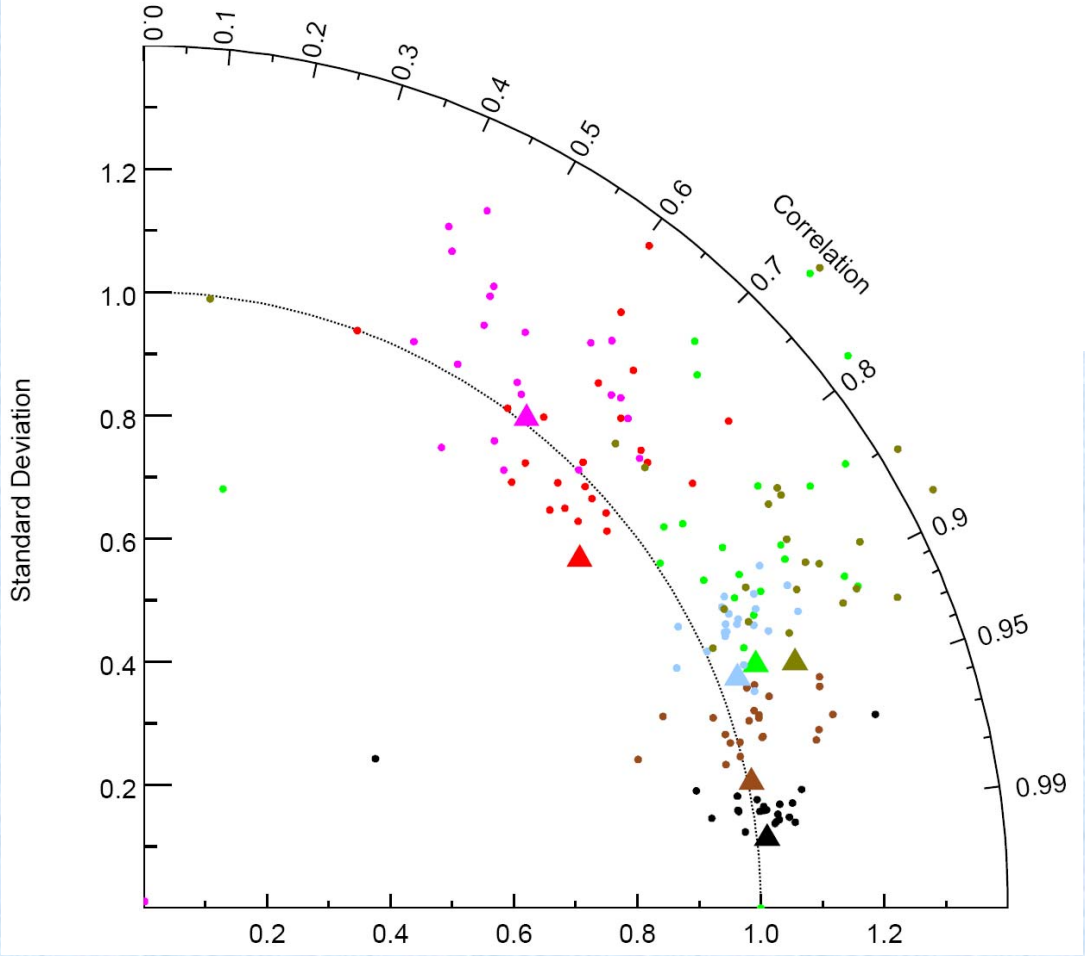
Fields	Reference / alternate
Zonal and meridional wind Temperature,Geopotential, 2m air temperature, 2m humidity and 10 winds	ERA40 / NCEP-NCAR reanalysis
TOA Radiative Fluxes: Outgoing Longwave (OLR), clear-sky fluxes	ERBE /CERES
Precipitable water	RSS / NVAP
Precipitation	CMAP / Xie-Arkin
Specific Humidity	AIRS/ ERA40
Total cloud cover	ISCCP-D2 / ISCCP-C2
Sea surface temperature (SST)	HadiSST / ERSST
Wind stress (ocean)	ERA40 / NCEP-NCAR
Ocean surface fluxes: latent and sensible (pattern only)	SOC / ERA40

Annual cycle performance metrics

- **Evaluate the climatology (1980-1999) of CMIP3 20th Century simulations with:**
  - → **~ 20 well-observed atmospheric variables**
  - → **Space-scale: global domain, coarse model grid (T42: 128x64)**
  - → **Time-scale: annual cycle**
- **Error statistics calculated by summing over all grid cells and the 12 climatological months**

PCMDI

Taylor diagram for CMIP3 annual cycle global climatology (1980-1999)



Sea Level Pressure: ERA40 reference
Total precipitation rate: CMAP reference
Total Cloud Cover: ISCCP reference
LW radiation TOA (OLR): CERES reference
Reflected TOA Shortwave: ERBE reference
Air Temperature (850 hPa): ERA40 reference
Zonal Wind (850 hPa): ERA40 reference

- Variable dependent skill
- Multi-model mean "superiority"

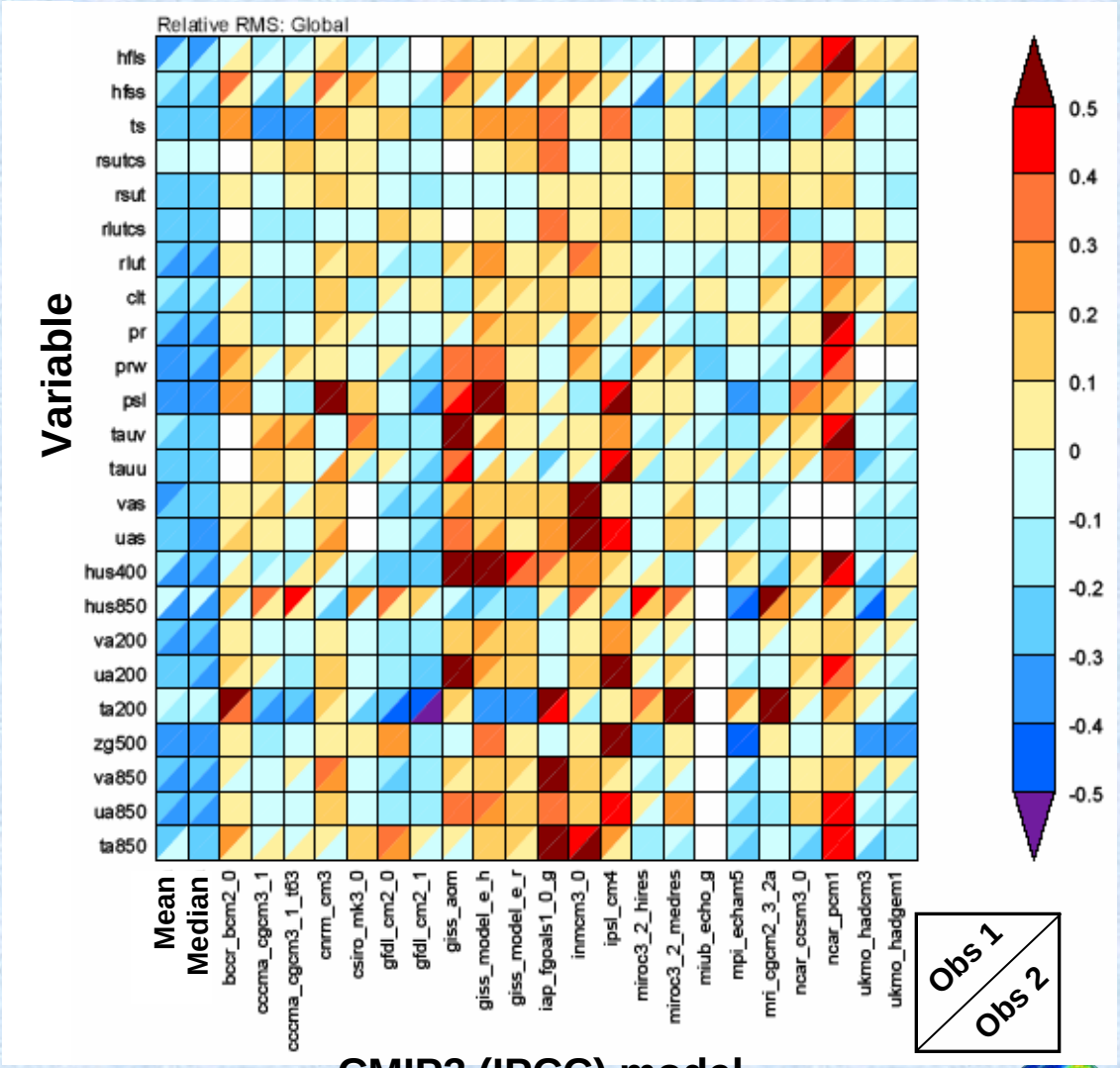
PCMDI

Annual cycle of global fields: Assessment of the relative skill (S) of individual CMIP3 models.

E_{vm} = RMS error in simulating the spatial pattern of the climatological annual cycle of variable v by model m

$$S_{vm} = \frac{E_{vm} - \hat{E}_v}{\hat{E}_v}$$

where $\hat{E}_v$ is the median of the individual error measures, E_{vm}

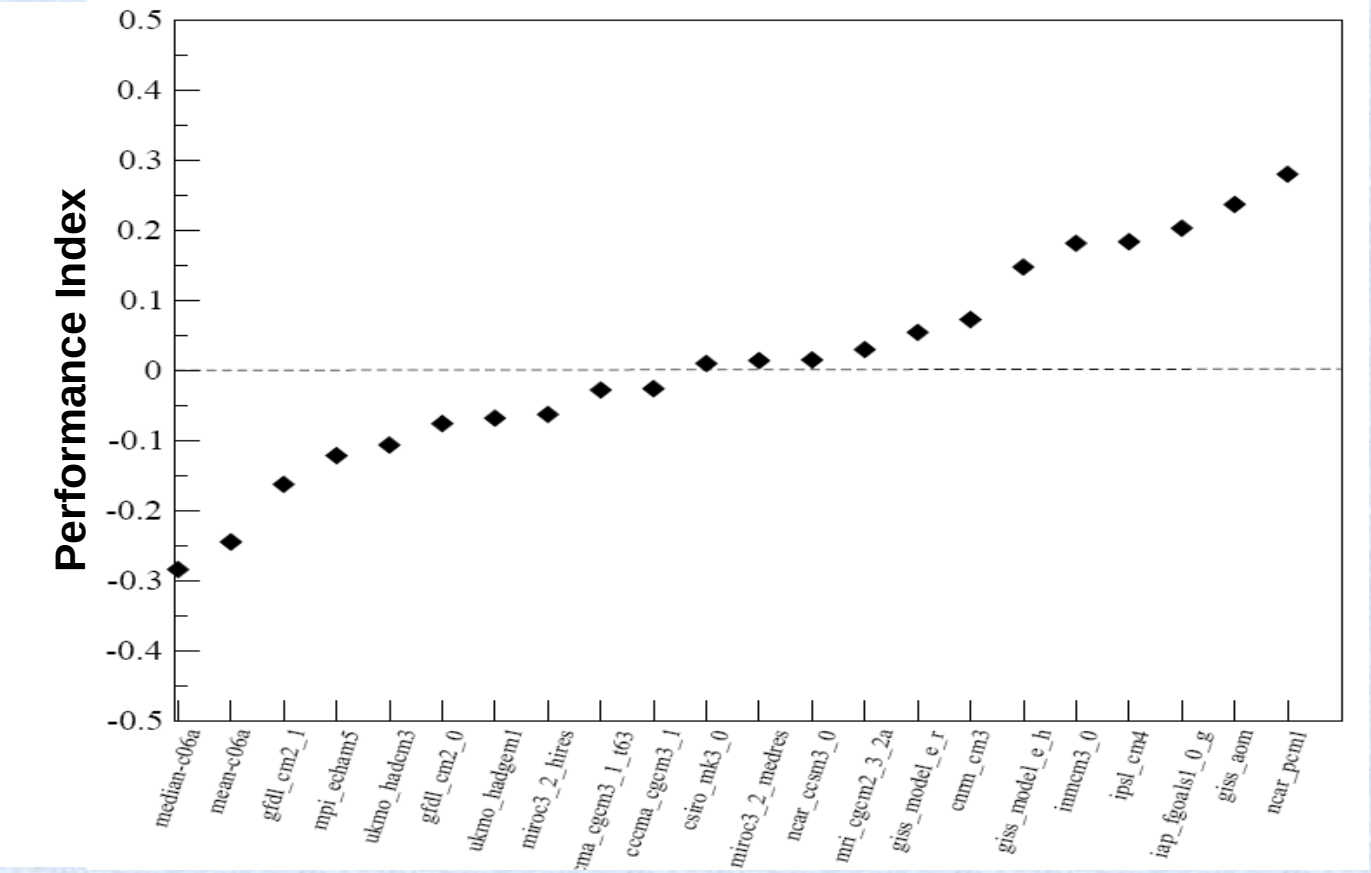


Gleckler, Taylor, & Doutriaux, 2008, *JGR, in press*

PCMDI

Exploring the value and limitations of a single "performance index".

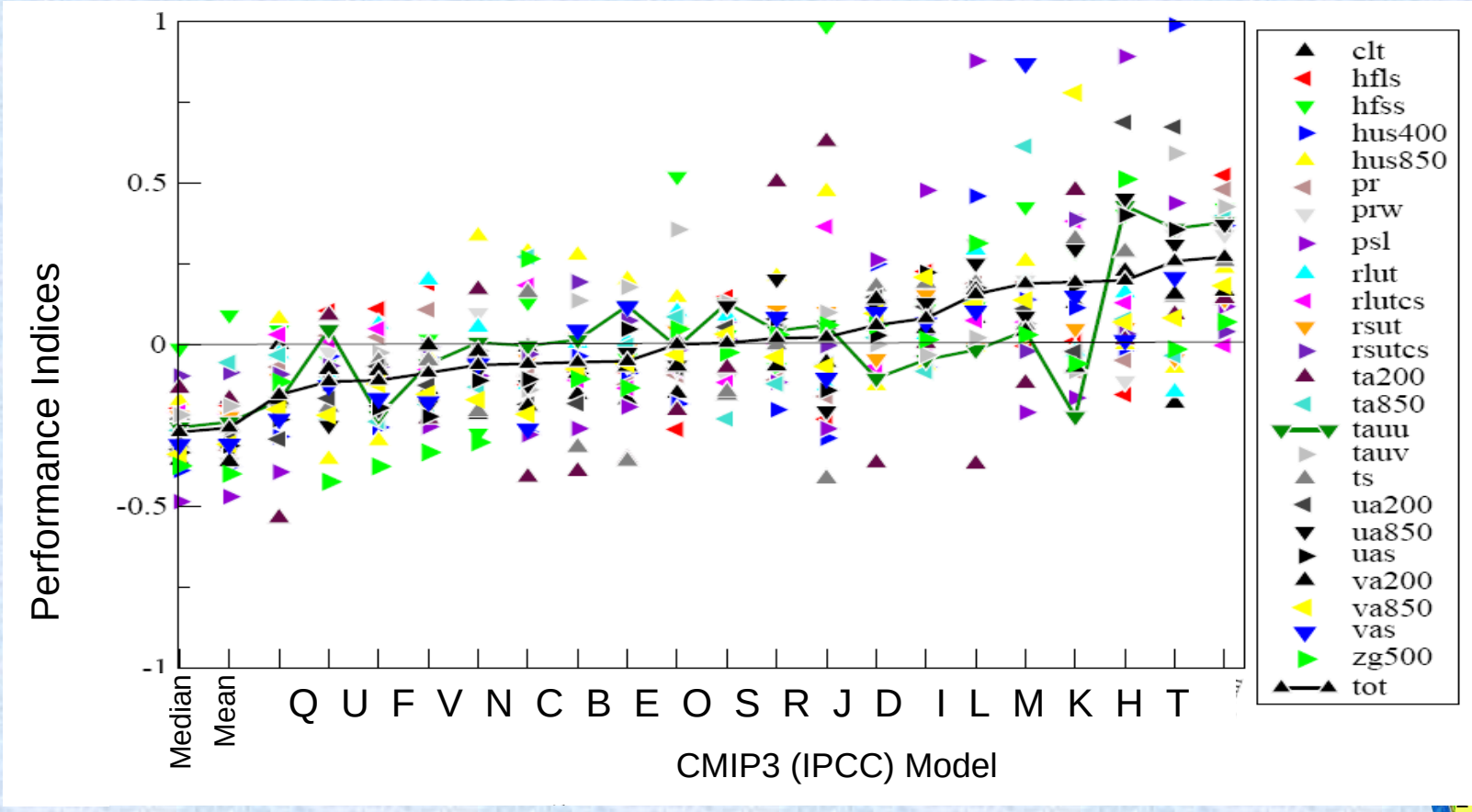
- **From performance portrait recall:** $$S_{vm} = \frac{E_{vm} - \hat{E}_v}{\hat{E}_v}$$
- **Let the "performance index" be the mean of S_{vm} over all the variables.**



PCMDI

Is the “performance index” meaningful/useful?

- **Little correlation between simulation of individual fields and an index.**
- **Ranking of models will depend on which metrics are included in index.**



PCMDI

Premature to unduly emphasize a single performance index

- **Fails to capture the complex error structure of models**
- **Depends on a number of factors (variable, region, time-scale, etc.)**
- **Invites simplistic interpretations of the relative value of specific models - the emphasis should be towards correct representation of the physics.**
- **Optimal weighting of different metrics contributing to a performance index likely depends on the application**

PCMDI

Do we know what is most important for reliable projections?

No, but …

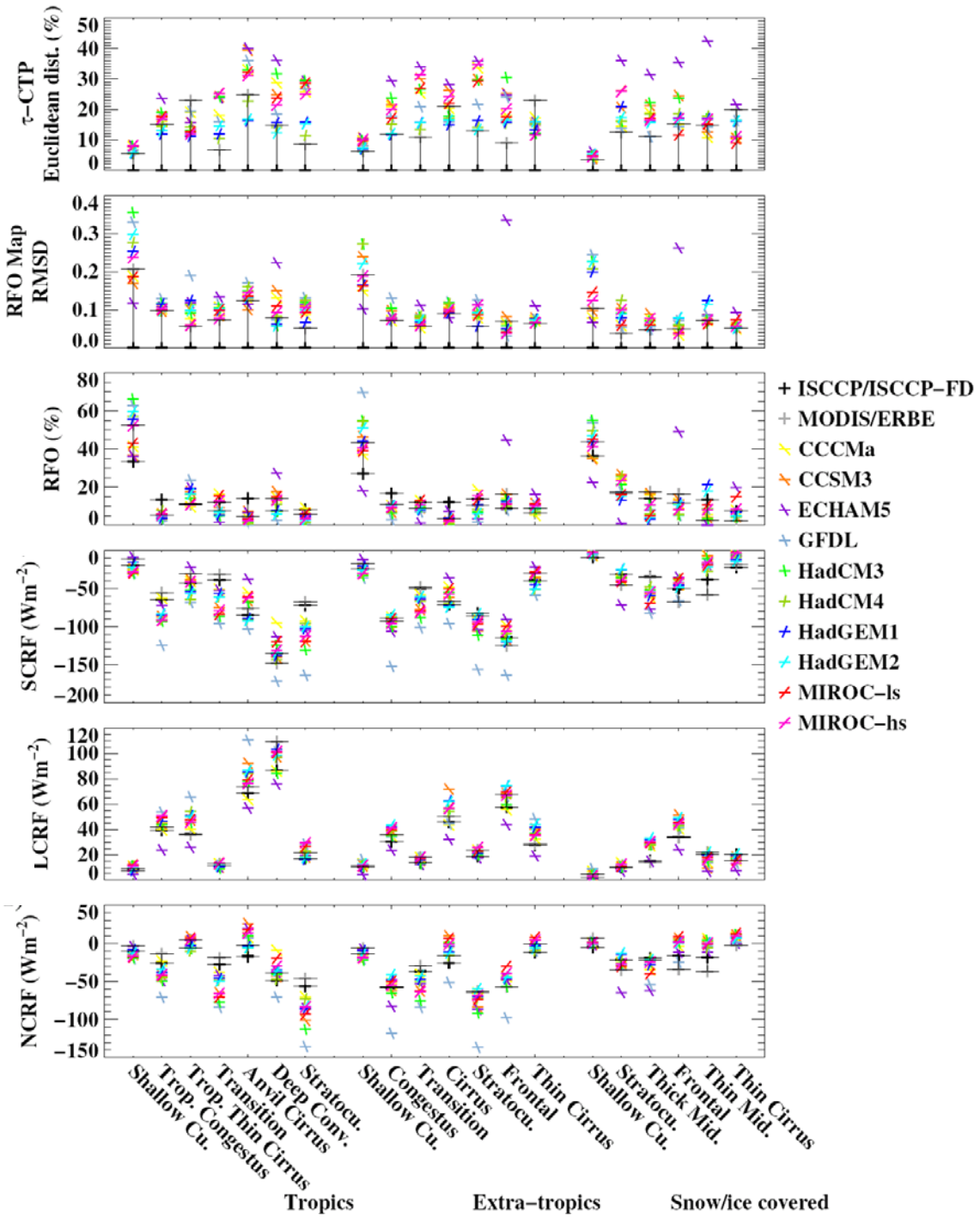
- Cloud-radiative effects are an obvious place to start
- Cloud-Feedback Model Intercomparison Project (CFMIP):

Objective of CFMIP-2 is to make an improved assessment of:

→ climate change cloud feedbacks by making progress in the
(1) evaluation of clouds simulated by climate models and the
(2) understanding of cloud-climate feedback processes.

→ From a practical standpoint – participating modeling groups provide "ISCCP simulator" output from standard experiments

PCMDI



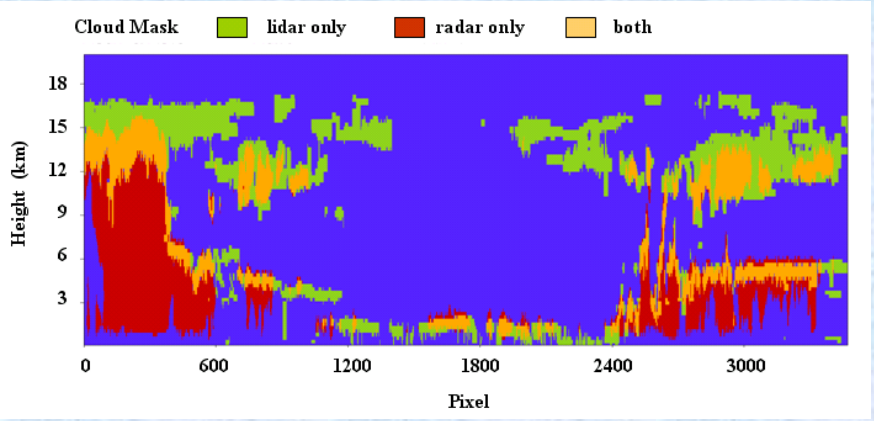
Williams and Webb, 2008

(Climate Dynamics, submitted)

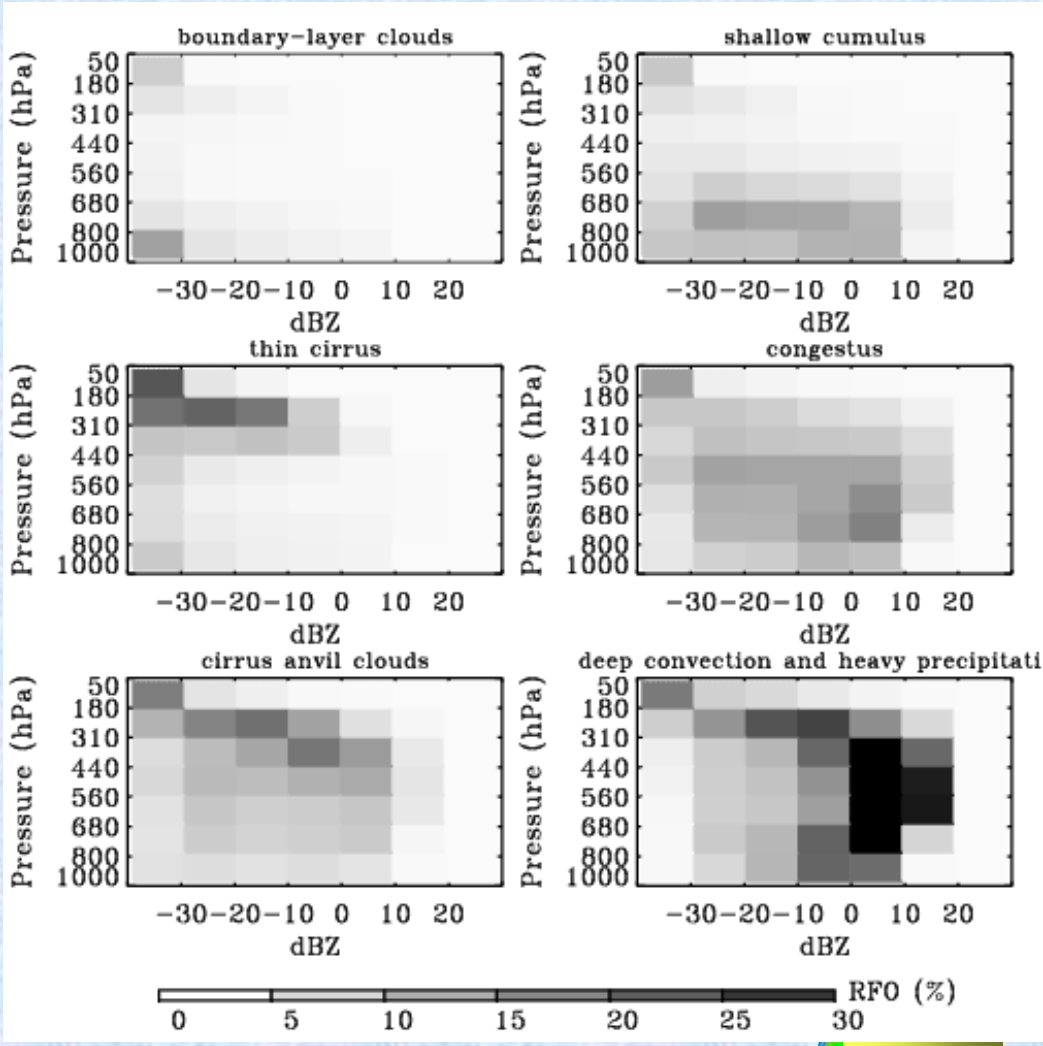
- CFMIP slab ocean exps
- Joint τ-CTP cloud amount histograms
- 5 yrs of daily mean ISCCP simulator and CRF
- Similar from MODIS

PCMDI

Define principal clusters of cloud regimes from observations (courtesy Yuying Zhang and Steve Klein of PCMDI)



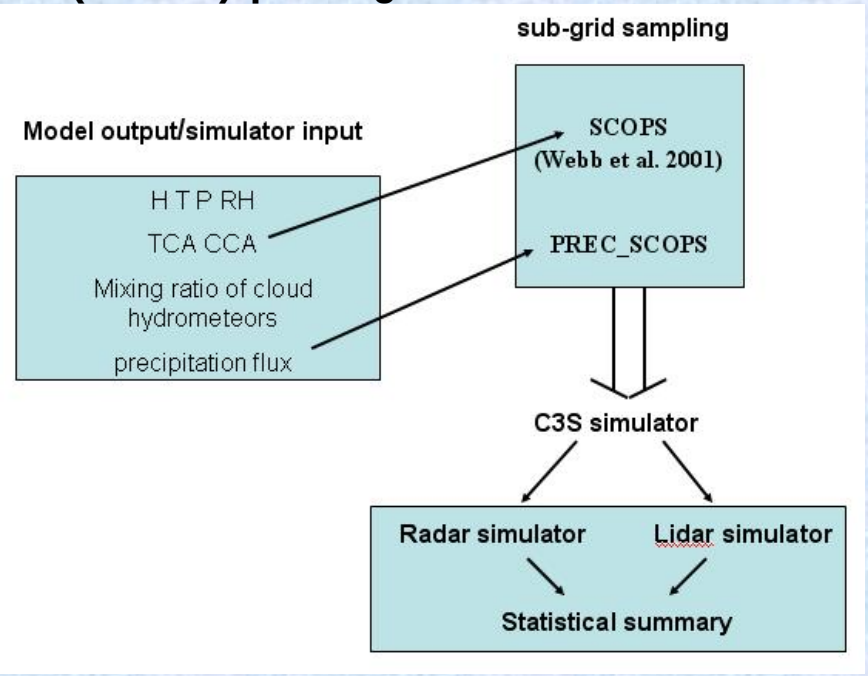
- Combined CloudSat and CALIPSO data provide most accurate description on vertical structure of cloud fields (Mace et al. 2007&2008)
- Patterns of cloud clusters defined using combined dataset (Zhang et al. 2007)



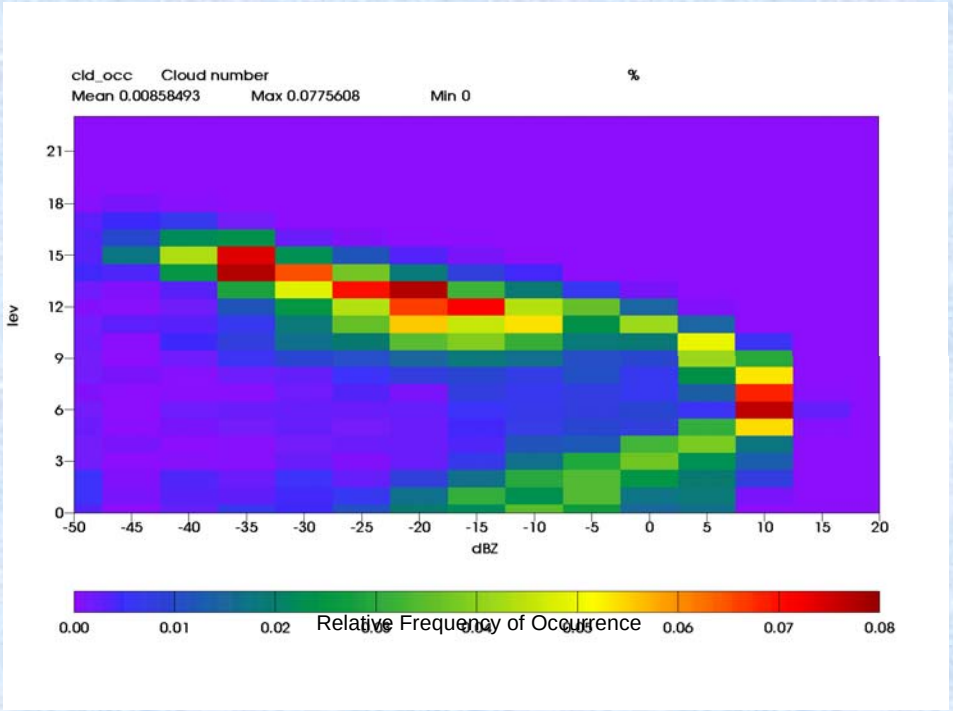
PCMDI

Goal: evaluate GCM simulations using combined radar and lidar data

Schematic of the CFMIP ISCCP/CloudSat/ CALIPSO simulator (CICCS) package



A sample: apply the radar simulator to the NCAR's CAM3 simulations



- **Development of CICCS is in collaboration with the Hadley Center and LMD (France), CSU, and UW**
- **Embed the CICCS in GCMs and produce the output similar to the observations**
- **Assess model performance using clustering analysis**

PCMDI

Beyond the mean climate . . .

- Variability also important simulating climate change
- Extensive diagnostic approaches exist
- Development of variability metrics in its infancy

PCMDI

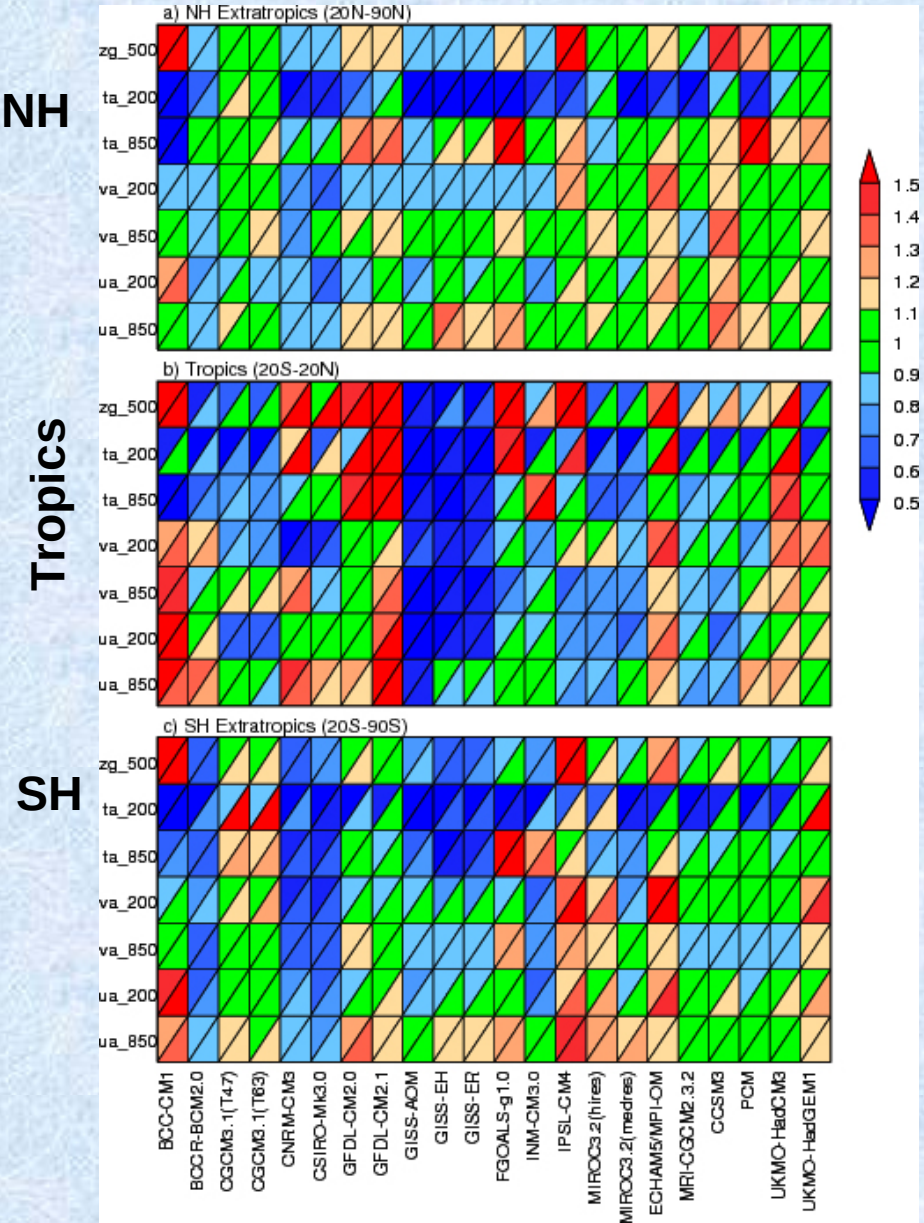
Monthly anomalies: Variance (model/reference)

- **Model anomaly amplitudes (domain average) relative to ERA40 and NCEP reanalysis (1980-1999):**

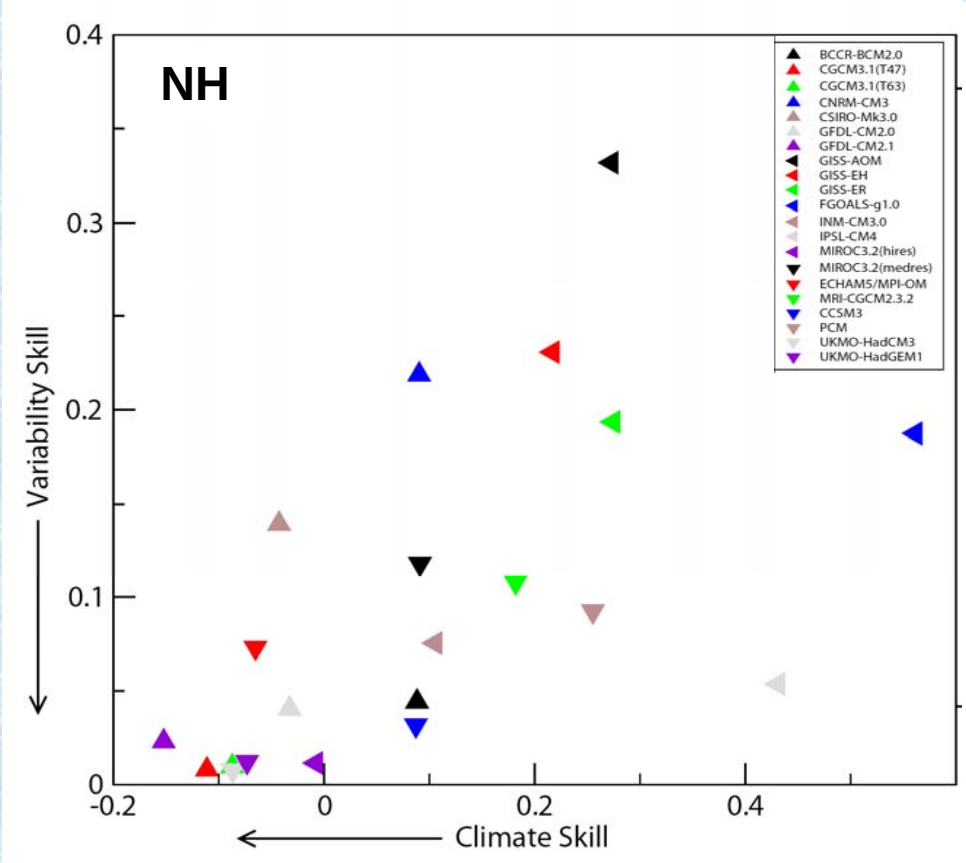
$$\frac{\text{Variance (model)}}{\text{Variance (reference)}}$$

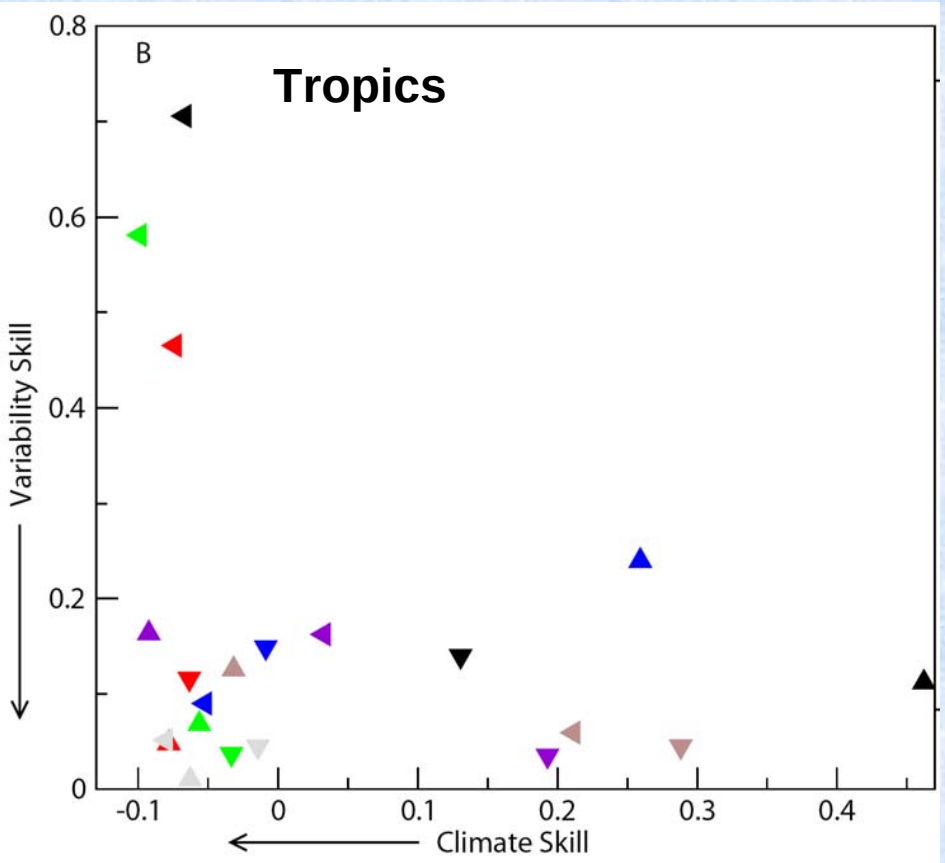
A "Model Variability Index":

$$MVI_{mr} = \sum_{f=1}^{F} \left[\beta_{mrf} - \frac{1}{\beta_{mrf}} \right]^2$$

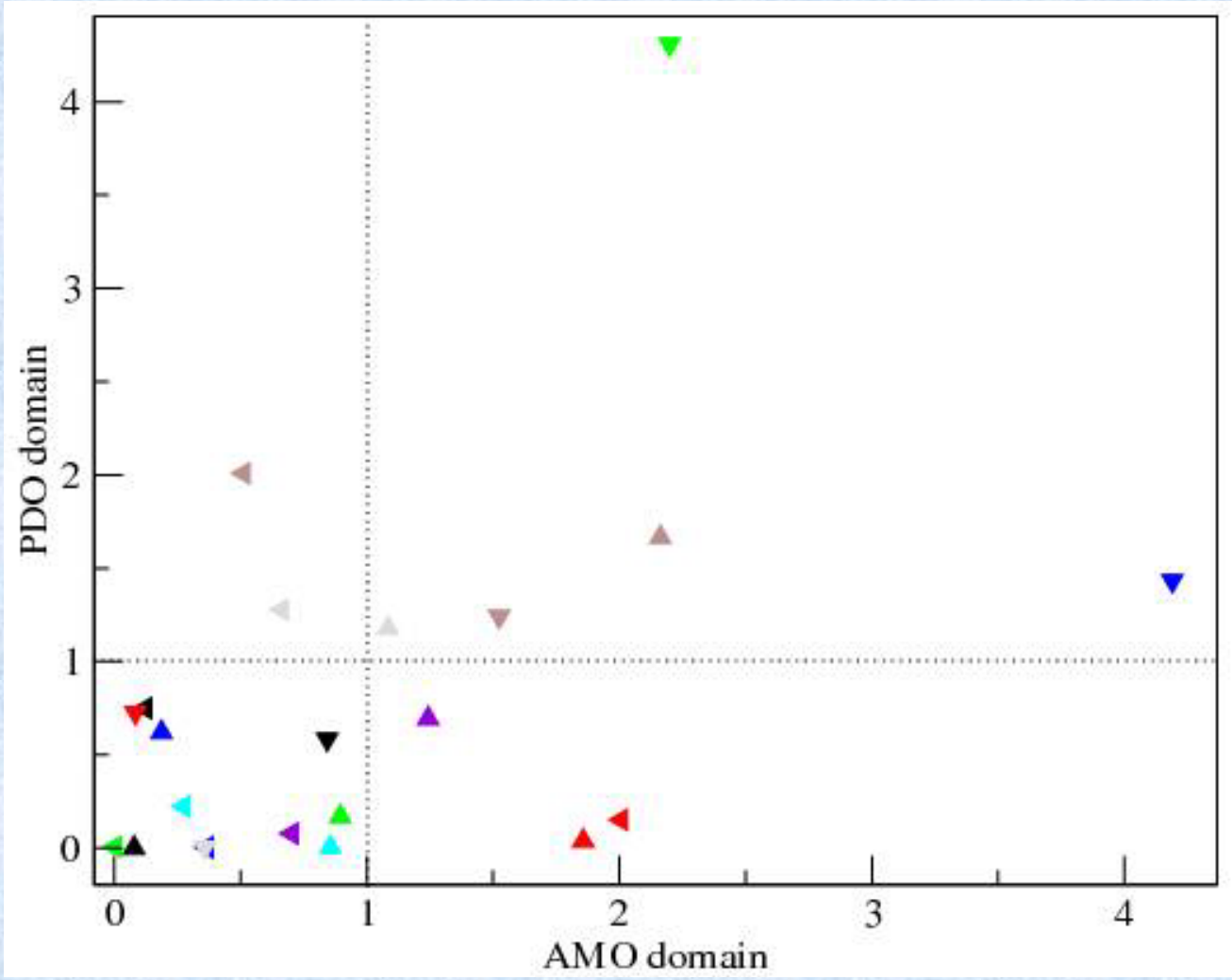


Model Skill: Mean climate vs. variability





SST anomalies: PDO vs AMO domains



PCMDI

Some conclusions: performance metrics gauging relative skill

- Mean climate and variability relative skill is regionally dependent
- Weak relationship between skill in simulating mean climate and variability
- Premature to unduly emphasize a single performance index - fails to capture the complex error structure of models
- Optimal weighting of different metrics contributing to a performance index likely depends on the application
- For the moment, ruling out models based on minimal requirements seems most justifiable

PCMDI

How have metrics helped us to date ?

- Force us to be more quantitative in our evaluation of models
- Enable us to track changes in model performance
- Help summarize the relative merits of different models
- Provide considerable evidence for the general superiority of the multi-model “mean simulation”

PCMDI

Looking ahead:

- Community working to develop a “basket” of metrics spanning a wide range of simulated processes and phenomenon
- Establish a minimum set of routine performance metrics, minimizing redundancy
- Explore relationships between skill in simulating present climate future projections
- Work towards scientifically justifiable strategies of weighting model results of future projections
- The more state-of-the-art observations to be incorporated into this work the better…

PCMDI